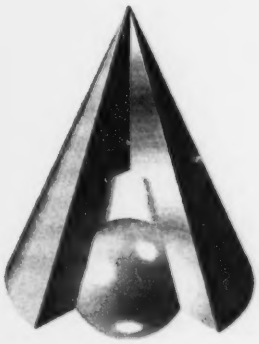
Catalogue no. 11F0019M — No. 340
ISSN 1205-9153
ISBN 978-1-100-20222-8

Research Paper

Analytical Studies Branch Research Paper Series

Immigrant Earnings Growth: Selection Bias or Real Progress?



by Garnett Picot and Patrizio Piraino

Social Analysis Division
24-I, R.H. Coats Building, 100 Tunney's Pasture Driveway
Ottawa, Ontario K1A 0T6

Telephone: 1-800-263-1136

Statistics Canada Statistique Canada

Canada

Immigrant Earnings Growth: Selection Bias or Real Progress?

by
Garnett Picot and Patrizio Piraino

11F0019M No. 340
ISSN 1205-9153
ISBN 978-1-100-20222-8

Statistics Canada
Social Analysis Division, Analysis Branch
24-I, R.H. Coats Building, 100 Tunney's Pasture Driveway
Ottawa, Ontario K1A 0T6

How to obtain more information:
National inquiries line: 1-800-263-1136
E-Mail inquiries: infostats@statcan.gc.ca

February 2012

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2012

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement. (http://www.statcan.gc.ca/reference/copyright-droit-auteur-eng.htm).

La version française de cette publication est disponible (n° 11F0019M au catalogue, n° 340).

Note of appreciation
Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, and Canada's businesses, governments, and other institutions. Accurate and timely statistical information could not be produced without the continued cooperation and goodwill of these partners.

Standards of service to the public
Statistics Canada is committed to serving its clients in a prompt, reliable, and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada, toll-free, at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca. Under "Our agency": click on "About us > The agency" and select "Providing services to Canadians."

Analytical Studies
Research Paper Series

The Analytical Studies Research Paper Series provides for the circulation, on a pre-publication basis, of research conducted by Branch staff, visiting fellows, and academic associates. The Analytical Studies Research Paper Series is intended to stimulate discussion on a variety of topics, including labour, business firm dynamics, pensions, agriculture, mortality, language, immigration, and statistical computing and simulation. Readers of the series are encouraged to contact the authors with comments and suggestions. A list of titles appears at the end of this document.

Papers in the Analytical Studies Research Paper Series are distributed to research institutes and specialty libraries. These papers can be accessed for free at www.statcan.gc.ca.

Publications Review Committee
Analysis Branch, Statistics Canada
24th Floor, R.H. Coats Building
Ottawa, Ontario K1A 0T6

Symbols

The following standard symbols are used in Statistics Canada publications:

.	not available for any reference period
..	not available for a specific reference period
...	not applicable
0	true zero or a value rounded to *0* (zero)
0^s	value rounded to *0* (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p}	preliminary
^{r}	revised
x	suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E}	use with caution
F	too unreliable to be published
*	significantly different from reference category ($p<0.05$)

Table of contents

Abstract **5**

Executive summary **6**

1 Introduction **8**

2 The issue: Using repeated cross-sectional data to estimate bias in immigrant earnings growth **9**

2.1 Empirical evidence from previous studies 11

3 Data **12**

4 Empirical results **15**

4.1 Regression estimates 17

5 Who is leaving the sample? **21**

6 Conclusion **25**

7 Appendix **26**

References **29**

Abstract

This paper studies the effect of selective attrition on estimates of immigrant earnings growth based on repeated cross-sectional data in Canada. Recent evidence from longitudinal data in the United States shows that the earnings gap between immigrants and the U.S.-born closes more slowly over time in the years following landing than previous cross-sectional estimates have suggested. This is because results based on repeated cross-sectional data contained a bias introduced by selective attrition of immigrants. This study uses longitudinal tax data linked to immigrant landing records in order to estimate the change in immigrant earnings and the immigrant–Canadian-born earnings gap. The results are compared with those from repeated cross-sectional data. When one focuses on the earnings growth of immigrants, earnings trajectories based on repeated cross-sections are found to be biased marginally upwards as a result of selective immigrant attrition. However, no evidence is found of a bias in the trajectory of the immigrant–Canadian-born earnings gap on the basis of repeated cross-sectional data in Canada. While low-earning immigrants are more likely than their high-earning counterparts to leave the cross-sectional samples over time, the same is true of the Canadian-born population. Thus, no evidence of a bias is observed when one compares immigrant earnings trajectories with the trajectories of the Canadian-born.

JEL classifications: J31, J61

Keywords: immigration, assimilation, longitudinal data, selection bias

Executive summary

The gap in earnings between immigrants and comparable native-born workers is perhaps the most studied topic in the economics of immigration. Such research would ideally be based on longitudinal data, tracking the earnings of immigrants after entry as they establish themselves in Canada. However, most existing Canadian research on immigrants' earnings trajectories is based on Census of Population cross-sectional data, not on longitudinal data. There are two reasons for this: longitudinal data have only recently become available, and they contain relatively few socio-economic covariates. Notably, in most Canadian longitudinal data, one cannot control for educational differences between immigrants and Canadian-born individuals.

Typically, researchers turn to repeated cross-sectional data from the Census to construct pseudo-longitudinal cohorts of individuals. Immigrants entering Canada during the 1991-to-1995 period are captured in the 1996 Census of Population one to five years after landing. Immigrants in this same cohort who remain in Canada are captured in the 2001 Census of Population six to ten years after landing, and so on. In this way, both immigrant earnings growth and the change in the immigrant–Canadian-born wage gap for various cohorts have been estimated.

However, the composition of pseudo-longitudinal cohorts changes over time as immigrants exit the host country or employment. This may introduce a bias in the earnings trajectories estimated from cross-sectional data. If, for example, immigrants who exit the sample are more likely to have poorer labour market outcomes than those who stay (and hence have an incentive to leave), the earnings trajectory based on pseudo-longitudinal cohorts will be biased upwards. As more time passes, any pseudo-longitudinal cohort will increasingly consist of "successful" immigrants from the original cohort—those with higher earnings. Hence, much of the increase in earnings over time (i.e., with years since landing) may result from a change in the composition of the cohort (a form of sample selection bias), not from a real increase in earnings. U.S. research on immigrant earnings has found such a bias.

This paper uses true longitudinal data from administrative records, linked to immigrant landing records. The earnings growth of immigrants and the earnings differential between immigrants and the Canadian-born are estimated over the years since migration for three immigrant cohorts that have arrived since the early 1980s. These data allow earnings outcomes to be estimated by using true longitudinal data and pseudo-longitudinal cohorts constructed from repeated cross-sections of the same data source. This approach eliminates differences in results that may stem from variation in collection modes and procedures across datasets. In order to more closely relate results to the existing literature, immigrant earnings trajectories with years since migration are also estimated by means of repeated cross-sections from the Canadian Census. The results from the two cross-sectional data sources (the Census and the administrative data) are then compared with those from the true longitudinal data source (the administrative data) in order to determine whether there is any evidence of a bias in the decline in the earnings gap between immigrants and Canadian-born individuals or in the earnings trajectories of immigrants alone.

The analysis provides little evidence of a bias in the immigrant–Canadian-born earnings gap computed from repeated cross-sections as compared to true longitudinal data. Although the lower-paid immigrants in the various cohorts are more likely to exit the sample than their higher-paid counterparts, the same appears to be true for the Canadian-born. That is, the earnings growth of both the immigrant cohorts and the Canadian-born cohorts is over-estimated in cross-sectional data, by roughly the same extent, hence the "gap" trajectory obtained by estimating the standard assimilation model on longitudinal data points to little bias in previous studies of earnings assimilation in Canada. This result is in contrast with at least some evidence from the United States. However, when one focuses on the earnings growth of immigrants and the Canadian-born separately, rather than on the earnings gap between them, the research suggests that earnings trajectories based on repeated cross-sections are biased upwards for both groups as a result of selective attrition.

1 Introduction

During their first few years in a host country, immigrants typically earn less than native-born individuals. However, immigrants' relative earnings rise in subsequent years as they obtain host-country experience, acquire language skills, and learn about local labour markets (Chiswick 1978; Meng 1987; Borjas 1999). For example, during their first five years in Canada, immigrants who landed during the late 1970s had earnings that were 85% of those of their Canadian-born counterparts; after 11 to 15 years, their relative earnings had reached 92%. The earnings of later immigrant cohorts were typically lower. The earnings of immigrants who landed in the early 1990s were 60% of those of their Canadian-born counterparts during their first five years in Canada, reaching 78% after 11 to 15 years (Frenette and Morissette 2003).[1]

Immigrants' earnings trajectories, and how these differ across successive landing cohorts, are best studied by means of longitudinal data comprising large sample sizes (to identify different landing cohorts) and information on socio-economic characteristics (to control for differences between immigrant and native-born populations). While some recent studies have used longitudinal administrative data (Hu 2000; Edin *et al.* 2000; Duleep and Dowhan 2002; Green and Worswick 2004; Lubotsky 2007; Aydemir and Robinson 2008), most research on immigrants' entry earnings and subsequent earnings trajectories is based on census data. This is because longitudinal data have only recently become available and contain relatively few socio-economic covariates. Notably, Canadian longitudinal data do not allow researchers to take into account educational differences between immigrants and the Canadian-born.

Typically, researchers turn to repeated cross-sections of the Canadian Census of Population in order to construct pseudo-longitudinal immigrant cohorts. For example, immigrants landing in Canada between 1991 and 1995 are captured in the 1996 Census one to give years after arrival. Immigrants in this cohort who remain in Canada will be captured in the 2001 Census six to ten years after landing, in the 2006 Census after 11 to 15 years after landing, and so on. On this basis, earnings growth and changes in the immigrant–Canadian-born wage gap have been estimated for various immigrant cohorts.

However, the samples in these pseudo-longitudinal cohort panels change over time, as some immigrants exit the host country. For Canada, Aydemir and Robinson (2008) focused on young male immigrants, a very mobile group, and estimated that about one-third left during their first twenty years in Canada, more than half of that group leaving in their first year. Exit rates among immigrant cohorts as a whole are likely lower. Immigrant exit may introduce a bias in the earnings trajectories estimated from cross-sectional data. If, for example, immigrants who exit tend to have poorer labour market outcomes than those who stay (and hence have an incentive to leave), the earnings trajectory based on pseudo-longitudinal cohorts constructed from cross-sectional data will be biased upwards. As more time passes, the pseudo-longitudinal cohort will increasingly consist of "successful" immigrants—those with higher earnings. Hence, much of the earnings gains with years since landing may result from a change in the composition of the cohort, a form of sample selection bias, not from a real increase in earnings.

This is exactly the result found by Hu (2000) and Lubotsky (2007) in the United States. Specifically, the earnings gap between immigrants and the native-born closed twice as fast when estimated by using pseudo-longitudinal cohorts constructed from repeated cross-sections of the U.S. Census as when estimated by using true longitudinal data. Lubotsky concluded that the higher probability of out-migration among low-wage immigrants systematically led past

1. More precisely, these numbers represent log earnings ratios (immigrant earnings to the earnings of the Canadian-born). A number of studies have looked at the decline in relative entry earnings for successive entering cohorts of immigrants in Canada (Bloom and Gunderson 1991; McDonald and Worswick 1998; Baker and Benjamin 1994; and Grant 1999). Picot and Sweetman (2005) offer a review.

researchers to overestimate the wage gains of immigrants remaining in the United States. These findings paint a less optimistic picture of the economic outcomes achieved over time by immigrants in the United States.

Contributors to the immigration policy debate in the United States often cite the Canadian experience as a system where high-skilled immigration is actively encouraged. Whether a point system similar to that used by Canada would improve the economic outcomes of immigrants to the United States remains an open question. However, establishing whether immigrant earnings growth in Canada is overestimated, as it appears to be in the United States, will help inform policymakers in both countries.

The data used in this study are described in detail in Section 3. Essentially, longitudinal data based on individuals' annual tax returns are linked to immigrant landing records. This yields a large representative sample of workers that includes both immigrants and Canadian-born. The earnings growth of immigrants and the earnings differential between immigrants and the Canadian-born are estimated over the years since migration for three immigrant cohorts that have arrived since the early 1980s. These data allow earnings outcomes to be estimated by using true longitudinal data and pseudo-longitudinal cohorts constructed from repeated cross-sections of the same data source. The fact that one can obtain both cross-sectional and longitudinal results from the same data source is important, as this eliminates differences in results that may stem from variation in collection modes and procedures across datasets. In order to more closely relate these results to the existing research literature, immigrant earnings trajectories are also estimated by means of pseudo-longitudinal cohorts constructed from repeated cross-sections of the Canadian Census.

The analysis provides little evidence of a bias in the immigrant–Canadian-born earnings gaps computed from repeated cross-sectional data. Although lower-paid immigrants in the three cohorts are more likely to exit the cross-sectional sample, the same is true of the Canadian-born. That is, the earnings growth of both the immigrant and the Canadian-born cohorts is over-estimated in cross-sectional data, by roughly the same extent. Hence, the earnings gap trajectory estimated by using a standard earnings model and longitudinal data indicates that there is little bias in previous studies of immigrants' relative earnings in Canada. This result is in contrast to the evidence for the United States.

The rest of the paper proceeds as follows. Section 2 explains the potential bias in immigrant earnings growth and reviews the small empirical literature on the issue. Section 3 describes the data used in this study, including the strengths and weaknesses of these data. The empirical findings are presented and discussed in Sections 4 and 5. Section 6 concludes.

2 The issue: Using repeated cross-sectional data to estimate bias in immigrant earnings growth

With respect to estimating immigrant earnings growth, the ideal situation would be one where all immigrants in successive landing cohorts remained in Canada over an extended period of time (for example, 20 years) and where longitudinal data on earnings and socio-demographic characteristics were available for each cohort. Producing unbiased estimates of earnings growth would be straightforward in this case. However, the situation is more complex. Some immigrants leave Canada, and, after an extended period of time, a subsample of 'long-term' immigrants remains. The composition of this subsample may be different from that of the initial landing cohort, particularly if the immigrants who left were those who had fared least well. Another consideration is that researchers may not have longitudinal data and may instead use repeated

cross-sectional data to construct pseudo-longitudinal immigrant cohorts. The issue of compositional bias again arises.

Focusing on a sample of long-term immigrants, this paper estimates earnings growth by using longitudinal data with a fixed sample composition as well as repeated cross-sectional data with a changing sample composition. By comparing results from the two approaches, one finds that the degree to which estimates based on repeated cross-sectional data are affected by compositional bias can be identified.

This paper does not address the bias resulting from the fact that the sample does not include all immigrants who entered Canada in any particular cohort but rather is restricted to long-term immigrants. However, it is the potential bias resulting from compositional changes within cohorts that is of primary concern to researchers.

Figure 1 explains the difference in the measurement of immigrant earnings growth in longitudinal and repeated cross-sectional data, as this difference applies to the Canadian Census. The rows of the figure indicate the year of arrival for three immigrant cohorts (1985, 1990, 1995), while the columns show the year in which earnings are measured. In each cell, $E(w)$ represents the average earnings measured at time c (column) for the cohort of immigrants who arrived in Canada at time r (row).

Panel A of Figure 1 shows that the cross-sectional samples will yield estimates of immigrant earnings growth that will depend on the rate of immigrant exits from the sample. For instance, the first row shows that earnings for the 1985 cohort will be measured in 1990, with average earnings calculated for the subset of immigrants still in the sample after five years. In 2005, average earnings for the same cohort will be calculated for the subset of immigrants still in the sample after 20 years. Looking across years, one notes that each immigrant will contribute to the estimated earnings growth rate of the landing cohort for as long as she/he is in the sample. If those who leave tend to be a non-random sub-sample of those who initially entered, a composition bias in the estimated earnings trajectory will occur.

In panel B of Figure 1 (longitudinal data), the sample is restricted to those immigrants who are captured in the last year of data. For each immigrant cohort, this allows average earnings to be estimated for immigrants who remained in the sample;[2] this yields an unbiased estimate of earnings growth for this subset of individuals.[3]

2. This does not exclude the possibility that in some years they may be absent.
3. This is not the same as estimating the earnings growth of the entering cohort had all individuals stayed until 2005. This estimation could be obtained from the longitudinal data on those who remained, only if one were willing to assume that attrition is based on permanent attributes that are not related to immigrant earnings growth over time. This interpretation is not attempted, as the focus of the paper is to test whether existing estimates of immigrant earnings growth obtained by using repeated census cross-sections are biased as a result of the changing composition of the immigrant cohort in the cross-sectional data. It is this issue that has concerned researchers using cross-sectional data and that on which most American research has focused (e.g., Lubotsky 2007).

Figure 1
Measures of average immigrant earnings, longitudinal versus cross-sectional data

	Year of observation			
Year of arrival	1990	1995	2000	2005
Panel A - Repeated cross-section				
1985	E(w| 5 years)	E(w|10 years)	E(w|15 years)	E(w|20 years)
1990	...	E(w| 5 years)	E(w|10 years)	E(w|15 years)
1995	...	...	E(w| 5 years)	E(w|10 years)
Panel B - Longitudinal data				
1985	E(w|20 years)	E(w|20 years)	E(w|20 years)	E(w|20 years)
1990	...	E(w|15 years)	E(w|15 years)	E(w|15 years)
1995	...	...	E(w|10 years)	E(w|10 years)

Note: *E(w)* represents the average earnings measured at time *c* (years shown in the columns) for the cohort of immigrants who arrived in Canada at time *r* (years shown in the rows).

Note that low-earning immigrants need not disproportionately emigrate from a country in order for this bias to arise and that data on emigration rates are not needed in order to estimate the bias. Instead, the main concern is disproportionate exit from employment by low-earning immigrants (rather than exit by emigration), and the sample of interest is the employed population. Indeed, when one estimates immigrant earnings trajectories by using repeated cross-sectional Census data, only observations with positive earnings in each cross-section are typically used.

2.1 Empirical evidence from previous studies

There is a very small literature that examines whether selective out-migration of immigrants results in a bias in cross-sectional estimates of immigrants' economic outcomes. On the basis of U.S. data, Hu (2000) and, more recently, Lubotsky (2007) concluded that selective emigration results in an overestimation of immigrants' economic outcomes.[4] In particular, Lubotsky used longitudinal earnings data from Social Security records for the 1951-to-1997 period and showed that the earnings gap between immigrants and the native-born closed half as fast in the longitudinal data as in repeated cross-sectional data from the decennial U.S. Census. As Lubotsky pointed out, however, the "bias" was not consistent across all entering cohorts, being most evident among the cohort that entered during the 1970s and only marginally evident among the cohorts that entered during the 1960s and 1980s.

Two papers utilize Canadian data to address this issue, although in a less direct manner than is done in the U.S. research, where the results from longitudinal and cross-sectional data are compared directly. Both Canadian papers are based on the Survey of Labour and Income Dynamics (SLID), a six-year longitudinal panel of Canadian workers in which immigrants can be identified. Hum and Simpson (2000) examined earnings growth over the 1993-to-1997 period and found that, even in the unadjusted longitudinal data, there was little change in the wage gap between immigrant and Canadian-born men. Earnings growth was about the same for both groups over the five-year study period. Among women, an increase, rather than a decrease, in the unadjusted wage gap was observed, as earnings growth was greater among Canadian-born than among immigrant women. Furthermore, employing a fixed-effects model, Hum and Simpson concluded that there was no evidence of a narrowing wage gap between immigrant

4. A similar conclusion is reached by Edin *et al.* (2000) in their analysis of Swedish data.

and Canadian-born men.[5] This is in contrast with virtually all existing Canadian studies based on repeated cross-sectional Census data. Hum and Simpson concluded that their results provide a warning that evidence from cross-sectional data, which may be prone to bias resulting from unobserved worker heterogeneity, should be interpreted cautiously.

In a more recent paper, Skuterud and Su (2009) pooled four panels of the SLID collected between 1993 and 2004 in order to augment the longitudinal sample of immigrants and Canadian-born. Contrary to Hum and Simpson (2000), they found evidence of considerable economic gains among immigrants. More relevant to this discussion, Skuterud and Su also attempted to address the issue of a bias in immigrant wage growth. Since the panels in their data are quite short, they utilized a substantially different approach than the one used in this paper or that used by Lubotsky (2007). Specifically, Skuterud and Su used a fixed-effects model to eliminate, to the extent possible, the effect of unobserved individual characteristics on both emigration and wage growth (i.e., the effects of selective out-migration on wage growth). They concluded that the fixed-effects approach changed the estimates of wage growth relatively little and did not imply substantially lower immigrant wage growth in longitudinal data as the US literature has tended to find (e.g., Lubotsky 2007). Their results point to the possibility that the nature of out-migration is different in Canada than the United States and that an upward bias in the existing cross-sectional estimates of immigrants' economic outcomes should not be expected.

By taking advantage of higher-quality administrative data, this study can help to shed light on these contrasting Canadian results. Moreover, given the longer reference periods being used, the effect of selective 'exits' can be examined and compared with results from the U.S. research. This is done by adopting the same analytical approach followed in Hu (2000) and Lubotsky (2007), which consists in selecting immigrant samples in terms of years since migration and examining earnings trajectories over this period.

3 Data

This study uses three data sources: the Longitudinal Administrative Database (LAD); the Longitudinal Immigration Database (IMDB); and the Census of Population. The LAD is a random, 20% subset of the T1 Family File (T1FF). *T1* refers to the *T1 General (Income Tax and Benefit Return)* for individuals. The T1FF is a yearly cross-sectional file of all individual tax filers in Canada and their families. Although one has to file an individual income tax return in order to be captured in the T1FF (and hence in the LAD), population coverage is very high, at around 95% for the working-age population. This is because tax rebates encourage individuals with little or no taxable income to file a return. An individual's records on the LAD are linked across years by means of a unique identification number; a longitudinal profile is thus created. In addition to annual earnings data, the LAD contains information on each individual's date of birth and sex.[6]

The IMDB consists of immigrant landing records and T1 personal income tax files. The former provide information on immigrant characteristics, while the latter provide longitudinal information on financial characteristics, employment earnings having greatest relevance to this study. Given the near-universal coverage of tax files, the IMDB allows the earnings trajectories of immigrant cohorts to be followed from the early 1980s up to 2005. In this paper, a linked LAD-IMDB data set is used. The linkage is possible as a result of the unique individual identifier available on both data sets.

5. This finding is confirmed in a successive study, which is based on the same dataset (Hum and Simpson 2004).
6. The definition of earnings includes wages, salaries, and commissions, before deductions, as well as taxable receipts from employment other than wages, salaries, and commissions (e.g., tips, gratuities, or directors' fees). It excludes self-employment income. More details on the dataset are available in Statistics Canada (2009).

The empirical analysis focuses on three successive immigrant cohorts: those who landed in Canada in 1985-to-1989, 1990-to-1994, 1995-to-1999. Given that data are available up to 2005, the earnings trajectories of the three cohorts can be followed for twenty, fifteen, and ten years after landing, respectively. Immigrant earnings over time are analyzed both in absolute terms and relative to Canadian-born individuals (i.e., the immigrant–Canadian-born earnings gap).

Like a number of previous studies, the analysis focuses on men in order to avoid complications arising from selective labour force participation. Immigrants are defined as foreign-born individuals who were 25 to 44 years of age at the time they landed in Canada, as reported on their landing record.[7] To generate earnings trajectories for Canadian-born men that match those of immigrants in the study, the Canadian-born comparison groups were formed from the same birth cohorts as those used for immigrants. Finally, the overall sample is restricted to person-year observations for ages 25 to 64.

A useful feature of the LAD-IMDB is that it allows both 'cross-sectional' and 'longitudinal' samples to be drawn. Because the data source is updated annually with new observations, the annual files remain cross-sectionally representative (Statistics Canada 2009). Cross-sectional samples of all person-year observations with positive earnings were selected for 1990, 1995, 2000, and 2005—all census years.[8] The pooled cross-sectional samples are then used to construct pseudo-longitudinal cohorts in a manner consistent which previous census-based studies of immigrants' earnings.

The longitudinal sample consists of annual earnings data for men in each immigrant cohort and each Canadian-born cohort for all years available. The crucial selection criterion for inclusion in the longitudinal sample is that individuals must have positive earnings in the latest year of data. This is the appropriate definition since the goal is to assess the bias in cross-sectional estimates of immigrant earnings growth. In fact, estimates of immigrant earnings assimilation from Census-pooled waves are based on a positive-earnings restriction in each cross-section used.

A pseudo-longitudinal sample is also constructed from the 1990, 1995, 2000, and 2005 cross-sectional Census files.[9] This allows estimates of earnings growth based on administrative and Census data to be compared. For consistency with the administrative sample, only males aged 25 to 64 with positive earnings are included.[10]

Several features of the data offer advantages over previous studies of immigrants' earnings trajectories, particularly in comparison with the Social Security earnings records used in the United States (Lubotsky 2007). First, the data used in this study do not result from a match of administrative records with survey data. This means that potential bias arising from non-random matches is not a consideration. Moreover, earnings trajectories can be compared by means of longitudinal and quasi-longitudinal samples drawn from the same data source; this avoids comparability issues that arise when different datasets are used. A second advantage is that the earnings data used here are not top-censored;[11] this avoids concerns related to top-coding of

7. Immigrants who were outside this age range at time of landing, as well as temporary foreign workers, are dropped from the analysis. The lower age limit is imposed because the labour market experience of very young immigrants is likely to be more similar to that of Canadian-born workers than to that of adult immigrants. The upper age limit serves to focus on immigrants with a longer window of time since migration. The sensitivity of the main results to this restriction is tested in the appendix.
8. The actual exclusion rule is *earnings>CAN$500*. Robustness checks were performed on various thresholds ($0, $1; $100; $1,000) with no effect on the paper's findings.
9. This study uses information from the Census "long" form, which was randomly administered to 20% of the population.
10. Also for the sake of consistency, this paper considers only immigrants who arrived in Canada as adults (between 25 and 44 years of age).
11. Top-censoring occurs when all earnings above a specified level, for example, $200,000, are reported as being at that level.

the sample and associated changes in the earnings ceiling over time. Finally, the data used in this study allow landed immigrants to be differentiated from temporary foreign workers.

Nonetheless, the LAD-IMDB has its shortcomings. Most notably, longitudinal earnings data are available only for individuals filing a tax return (although this represents about 95% of the working-age population). When no tax return is observed, it is not possible to determine whether this is because individuals are not employed, have left Canada, or have not filed a tax return. While the LAD-IMDB contains some information on immigrants' characteristics at the time they landed in Canada, such as educational attainment and intended occupation, this database does not contain such information for Canadian-born individuals. Hence, estimates of earnings differentials may be hampered by this lack of information. This shortcoming does not affect the analysis presented in this paper, however, for reasons outlined later in the paper. Finally, the IMDB identifies only immigrants who have landed since 1980. Foreign-born individuals who arrived in Canada before 1980 are included in the file but cannot be "flagged" as immigrants. Hence, the comparison group includes not only individuals born in Canada, but also immigrants who arrived prior to 1980. While this does not affect estimates of immigrants' absolute earnings, estimates of the immigrant–Canadian-born earnings gap will include "long-duration" immigrants in the comparison group.

The magnitude of this problem and the comparability of the LAD-IMDB and the Census of Population are assessed in Appendix Table 6. To make the comparison possible, the Canadian-born group in the Census is "augmented" by immigrants who landed before 1980.[12] The results for the LAD-IMDB sample and the census sample show that the two data sources are quite consistent in terms of absolute cohort size and the share of cohorts comprising immigrants. In the Census comparison group, Census information is used to determine the share of the comparison group consisting of longer-term *immigrants* as opposed to individuals born in Canada. For the 1990-to-1994 cohort, just over 3% of the comparison group are longer-term immigrants (in this case, those who have been in Canada for sixteen years or more), and over 96% are Canadian-born.[13] The share is higher in the earlier cohort: just under 6% of the comparison group for the 1985-to-1989 cohort consists of immigrants who landed in Canada before 1980. For the latest cohort, the share of immigrants in the comparison group is negligible.

Given the relatively small shares of the comparison group who are longer-term immigrants and their economic resemblance to the Canadian-born population, this comparison-group issue is not seen as being particularly troublesome.[14] As well, the fact that the extent of the "contamination" varies from cohort to cohort is not of concern, since only within-cohort comparisons of the earnings gap trajectories based on longitudinal and cross-sectional data are of interest. Finally, and perhaps most importantly, in the empirical analysis, one can compare the cross-sectional results based on the LAD-IMDB samples with the estimates obtained from the Census. In the next section, it will be shown that the Census results, which are not affected by this comparison group issue, are consistent with the estimates from the administrative data.

12. This is done only in order to obtain the descriptive statistics in Appendix Table 6. In the empirical analysis that follows, Census samples do not include immigrants landed before 1980.
13. Note that the Canadian-born group also includes child migrants (under 18 years of age) who arrived in Canada before 1980.
14. Furthermore, most of the attrition among immigrants takes place in the first few years in Canada (Aydemir and Robinson 2008). Therefore, it is unlikely that a small 'contamination' by longer-term immigrants would significantly affect the attrition probabilities among the comparison group.

4 Empirical results

The empirical results begin with some descriptive patterns based on unadjusted (raw) data. Tables 1 and 2 compare the level of immigrant earnings and the immigrant–Canadian-born earnings gap by years since migration for three different cohorts, by means of three different samples. The three samples are the following: (*i*) the longitudinal sample from the LAD-IMDB; (*ii*) the cross-sectional sample from the LAD-IMDB; and (*iii*) the cross-sectional sample from the Census. In essence, Tables 1 and 2 fill in the information outlined in Figure 1.

The differences between the longitudinal and cross-sectional samples from the LAD-IMDB (top two panels in Table 1) are the primary focus. For all cohorts, immigrant earnings during the first few years in Canada tend to be marginally higher in the longitudinal sample than in the cross-sectional sample. By the end of the study period (e.g., after 20 years in Canada for the 1985-to-1989 cohort), earnings are identical in the two samples. This is by design since the two samples themselves are identical by that time, consisting of all immigrants who were still in Canada and employed after 20 years. However, since the mean earnings are somewhat lower at the beginning of the period and identical by the end, the earnings growth is marginally steeper in the cross-sectional sample than in the longitudinal sample. This is what one might have expected to see on the basis of the discussion above. In terms of earnings gaps between immigrants and the Canadian-born, however, there is little variation between the longitudinal and cross-sectional samples. These patterns anticipate the major finding in the econometric analysis that follows. While there appears to be some differences in the absolute earnings growth of immigrants between the cross-sectional and the longitudinal samples as a result of selective attrition in the cross-sectional immigrant data, the same pattern of selective attrition is observed among Canadian-born workers. As a result, the bias stemming from this attrition is cancelled out when one compares earnings trends of immigrants to those of the Canadian-born. Hence, the earnings gap between immigrants and the Canadian-born closes over time at a similar pace in the two samples.

The bottom panel of Table 1 reports the same statistics for the samples drawn from the Census of Population. When one compares the pseudo-longitudinal cohorts constructed from the Census of Population and from the LAD-IMDB cross-sections, earnings trajectories are very similar. For all cohorts, immigrant earnings growth is virtually the same, and there are only minor differences in the decline in the earnings gaps for all cohorts (see Table 2). Average immigrant earnings tend to be higher in the Census than in the LAD-IMDB cross-sections. This is consistent with Frenette, Green, and Picot (2006), who found that income values in the bottom half of the distribution are lower in the tax data than in the Census. All in all, the cross-sectional samples from the two data sources yield similar immigrant earnings trajectories, in both absolute and relative (to Canadian-born) terms.

Table 1
Average immigrant earnings, longitudinal versus cross-sectional data

	Year earnings are measured			
	1990	1995	2000	2005
Longitudinal data in LAD-IMDB				
1985-to-1989 cohort				
Sample size (number)	12,080	11,570	12,173	14,075
Earnings (dollars)	32,100	38,200	48,600	49,100
Earnings gap (percent)	-30.2	-22.8	-15.8	-14.8
1990-to-1994 cohort				
Sample size (number)	...	19,832	21,496	24,934
Earnings (dollars)	...	27,200	40,700	43,000
Earnings gap (percent)	...	-40.4	-26.7	-25.6
1995-to-1999 cohort				
Sample size (number)	...	...	21,278	24,902
Earnings (dollars)	...	...	35,900	43,500
Earnings gap (percent)	...	...	-30.0	-22.2
Repeated cross-section in LAD-IMDB				
1985-to-1989 cohort				
Sample size (number)	19,484	15,567	14,637	14,075
Earnings (dollars)	31,600	36,600	46,300	49,100
Earnings gap (percent)	-29.0	-23.6	-16.6	-14.8
1990-to-1994 cohort				
Sample size (number)	...	29,032	26,697	24,934
Earnings (dollars)	...	26,100	38,950	43,000
Earnings gap (percent)	...	-40.5	-27.6	-25.7
1995-to-1999 cohort				
Sample size (number)	...	...	28,089	24,902
Earnings (dollars)	...	...	34,100	43,500
Earnings gap (percent)	...	...	-30.7	-22.1
Census				
1985-to-1989 cohort				
Sample size (number)	20,746	17,625	16,645	16,493
Earnings (dollars)	32,700	37,300	45,000	50,200
Earnings gap (percent)	-24.2	-18.1	-12.1	-9.3
1990-to-1994 cohort				
Sample size (number)	...	28,483	26,326	25,354
Earnings (dollars)	...	27,600	39,100	45,300
Earnings gap (percent)	...	-34.4	-21.4	-18.4
1995-to-1999 cohort				
Sample size (number)	...	...	27,819	25,643
Earnings (dollars)	...	...	36,000	45,200
Earnings gap (percent)	...	...	-22.4	-16.2

Notes: Authors' calculations from the LAD-IMDB and the Census of Population. The sample size refers to immigrants only. In each year, the population consists of males 25 to 64 years of age with positive earnings. Immigrants migrated between 25 and 44 years of age. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

Table 2
Immigrant earnings growth and decline in earnings gap, longitudinal versus cross-sectional data

Databases and cohorts	Earnings growth			Decline in earnings gap		
	From 1990 to 2005	From 1995 to 2005	From 2000 to 2005	From 1990 to 2005	From 1995 to 2005	From 2000 to 2005
	percent			percentage points		
Longitudinal data in LAD-IMDB						
1985-to-1989 cohort	53	...	...	15.4	...	...
1990-to-1994 cohort	...	58	...	...	14.8	...
1995-to-1998 cohort	...	...	21	...	...	7.8
Repeated cross-section in LAD-IMDB						
1985-to-1989 cohort	55	...	...	14.2	...	...
1990-to-1994 cohort	...	64	...	...	14.8	...
1995-to-1998 cohort	...	...	27	...	...	8.6
Census						
1985-to-1989 cohort	53	...	...	14.9	...	...
1990-to-1994 cohort	...	64	...	...	15.9	...
1995-to-1998 cohort	...	...	26	...	...	6.2

Notes: Author's calculations from the LAD-IMDB and the Census of Population. In each year, the population consists of males 25 to 64 years of age with positive earnings. Immigrants migrated between 25 and 44 years of age. The earnings gap is the difference between the earnings of immigrants and the earnings of the Canadian-born, divided by the earnings of the Canadian-born. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

4.1 Regression estimates

The patterns above are based on the unadjusted (raw) data, but most of the results reported in the literature are from some form of regression model. A standard econometric framework to examine the absolute and relative earnings of immigrants in longitudinal and repeated cross-sectional data is used. In order not to impose an assumption of constant earnings growth across cohorts, data across cohorts is not pooled; the earnings trajectories of successive immigrant cohorts are analyzed separately.

Evidence from the United States suggests that the wage progression of immigrants in the longitudinal data is not consistent across entering cohorts (Duleep and Regets 1997; Duleep and Dowhan 2002). Much of the estimated 'out-migration' bias in Lubotsky (2007) seems to derive only from the 1970-to-1979 cohort, not from the other two cohorts examined. The results from the unadjusted data above suggest there may be some cross-cohort differences in the Canadian data as well.

While most of the empirical literature focuses on the relative earnings growth of immigrants, it is useful to describe the trajectories in immigrants' earnings levels, as well as the earnings gap between immigrants and the Canadian-born, and to assess how the earnings growth and earnings gap differ in the cross-sectional and longitudinal data. Hence, the analysis starts by estimating the absolute earnings trajectories of entering immigrants, running separate regressions for each of the three cohorts. A simple way to capture these trends is to estimate the following regression for the entering cohorts:

$$w_{it} = \alpha + \beta_1 Age_{it} + \theta ysm_t + \varepsilon_{it} \qquad (1)$$

where: w_{it} is the log of annual earnings for individual i in year t; Age_{it} is a polynomial in the individual's age; and ysm_t is the number of years has spent in the host country since landing, which is specified as a categorical variable (0-to-5 years; 6-to-10 years; 11-to-15 years; and 16-to-20 years). In this immigrants-only regression, collinearity does not allow one to estimate period effects; hence, calendar time controls are not included. Table 3 reports the estimated coefficients for years since migration in model (1) for the three immigrant cohorts. Estimates are provided separately for the cross-sectional and longitudinal LAD-IMDB samples as well as for the Census sample.

For all three cohorts, there is evidence that the earnings trajectory is overestimated in the cross-sectional sample as compared to the longitudinal sample. For the 1985-to-1989 cohort, Table 3 shows that the earnings growth between 0-to-5 years and 16-to-20 years in Canada was 0.27 (about 31%) in the longitudinal LAD-IMDB and 0.33 (39%) in the cross-sectional LAD-IMDB. That is, immigrant earnings growth after 16-to-20 years in Canada differs by 8 percentage points. These results suggest an upward bias in the cross-sectional results. For the 1990-to-1994 cohort, the earnings growth after 11 to 15 years in Canada is 0.39 in the longitudinal LAD-IMDB and 0.49 in the cross-sectional LAD-IMDB, a 15-percentage-point difference. A bias is also observed with respect to the latest cohort (0.21 versus 0.27). Note also that the estimates from the cross-sectional LAD-IMDB sample are very much in line with those based on the Census. This confirms that the administrative data provide estimates of immigrants' earning trajectories similar to published results based on Census data.

Table 3
Immigrant earnings growth in Canada, longitudinal versus cross-sectional data

Years since migration	LAD-IMDB		Census
	Longitudinal	Cross-sectional	
	Column 1	Column 2	Column 3
	coefficient from regression model		
1985-to-1989 cohort			
6 to 10 years	0.117	0.065	0.057
11 to 15 years	0.271	0.266	0.237
16 to 20 years	0.271	0.333	0.314
1990-to-1994 cohort			
6 to 10 years	0.321	0.371	0.374
11 to 15 years	0.391	0.494	0.498
1995-to-1999 cohort			
6 to 10 years	0.216	0.275	0.267

Notes: The coefficients provide a rough estimate of percentage growth; e.g., the earnings growth in the longitudinal data between 0-to-5 years and 6-to-10 years in Canada for the cohort entering between 1985 and 1989 was approximately 11.7%. The reference category is immigrants with one to five years since migration. All coefficients are statistically significant at the 1% level. The table reports the coefficients on the "years since migration" dummy variables in model 1. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database, and Census of Population.

Many standard earnings regression models take into account education; hence, educational attainment at time of landing is added to equation (1). As a result of doing so, part of the bias between the longitudinal and cross-sectional samples would be eliminated, *if* that bias were driven by a higher probability of exit among less educated immigrants. However, this is not the case. The difference in immigrant earnings growth estimated from the two LAD-IMDB samples

is much the same whether or not education is taken into account (Tables 3 and 4).[15] However, looking at earnings growth over time *within* cohorts, one finds evidence of larger increases in the cross-sectional sample than in the longitudinal sample, especially for the 1990-to-1994 cohort. This provides indirect evidence of a higher probability of exit by low-earning immigrants within education groups.

From the estimates in Tables 3 and 4, one can infer that, among immigrants in the three cohorts analyzed, those exiting the sample were more likely to have lower earnings than those who stayed. As a result, changes over time in the composition of the repeated cross-sections resulting from selective exits among lower and higher earners introduced a bias in the absolute earnings trajectories estimated on cross-sectional data.

Table 4
Immigrant earnings growth in Canada (with controls for education), longitudinal versus cross-sectional data

Years since migration	LAD-IMDB	
	Longitudinal	Cross-sectional
	coefficient from regression model	
1985-to-1989 cohort		
6 to 10 years	0.143	0.096
11 to 15 years	0.320	0.325
16 to 20 years	0.335	0.408
1990-to-1994 cohort		
6 to 10 years	0.345	0.399
11 to 15 years	0.435	0.536
1995-to-1999 cohort		
6 to 10 years	0.243	0.287

Notes: The coefficients provide a rough estimate of percentage growth; e.g., the earnings growth in the longitudinal data between 0-to-5 years and 6-to-10 years in Canada for the cohort entering between 1985 and 1989 was approximately 14.3%. The reference category is immigrants with one to five years since migration. All coefficients are statistically significant at the 5% level. The table reports the coefficients on the "years since migration" dummy variables. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.
Sources: Statistics Canada, Longitudinal Administrative Database and Longitudinal Immigration Database.

However, most research does not focus on the earnings trajectories of immigrants alone; rather, it examines changes over time in the earnings gap between immigrants and Canadian-born individuals. To test for such bias, the Canadian-born comparison groups described in Section 3 are added to the analysis. To evaluate immigrants' earnings gains (with years since landing) relative to the comparison group, a standard empirical framework for this type of analysis is applied (Chiswick 1978; Borjas 1999).

Consider the following regression model of log annual earnings:

$$w_{it} = \alpha + \beta_1 Age_{it} + \beta_2 Year_{it} + \lambda I_i + \gamma M_i \cdot I_i + \theta ysm_{it} \cdot I_i + \varepsilon_{it} \qquad (2)$$

where the additional variables beyond those in equation (1) are the following: a vector of calendar-time dummies, $Year_{it}$; the immigrant's age at arrival in the host country, M_i, to proxy for foreign labour market experience; and a dummy variable identifying immigrant and native-born

15. The conditional regression on the Census sample is not run in this study, as the education categories used in the LAD-IMDB do not match those reported in the Census.

status, I_i. Note that, in model (2), the variables M_i (immigrants' age at arrival in the host country) and ysm_{it} (years since migrations) are now interacted with the immigrant-status dummy, allowing the earnings trajectory to differ between immigrants and the native-born. As before, a separate model is run for each of the three cohorts.

The coefficient on years since migration, θ, is the parameter of interest and measures the change in the immigrant–Canadian-born earnings gap with years spent in Canada, or, put another way, the rate of earnings convergence over time between immigrants and the Canadian-born. Immigrants' earnings gap at the time of entry and the effect of foreign experience on earnings in the host country are captured by λ and γ, respectively. In order to separately identify the coefficients on the variables *Age*, *Year*, *M*, and *ysm*, the restriction that the age and period effects, β_1 and β_2, be the same for immigrants and Canadian-born must be imposed. As explained in Borjas (1999) and Lubotsky (2007), this assumption is not trouble-free. However, most estimates of immigrant earnings growth are based on this assumption, and it is kept for the purpose of evaluating the difference in earnings growth measured by means of longitudinal and cross-sectional samples. The model is the same as that used by Lubotsky (2007) in the United States, against which these results can be compared.[16]

Columns 1 and 2 in Table 5 report the results from the estimation of equation (2) based on the longitudinal LAD-IMDB sample (column 1) and the cross-sectional LAD-IMDB sample (column 2). Column 3 shows the results from the Census sample. This analysis starts by comparing the results based on the two LAD-IMDB samples, thereby eliminating any differences due to data sources (survey versus administrative). For the 1985-to-1989 cohort, there are no major differences to speak of, either in the entry earnings gap or in the change in the gap over time. In the two samples, the gap at entry is around 0.33 to 0.35; after 16-to-20 years in Canada, it has been reduced by about 0.24 (roughly 27%). For the 1990-to-1994 cohort, the gap at entry is marginally larger among the cross-sectional sample (as one might expect if there were a bias). However, the difference is statistically insignificant, and there are no significant differences in the change in the gap between the samples; this is this study's main interest. The same applies to the 1995-to-1999 cohort. Overall, the results from the administrative data do not suggest the existence of a bias in repeated cross-section estimates of earnings assimilation—contrary to what was observed for absolute earnings. A similar conclusion is reached when the longitudinal results from the LAD-IMDB are compared with the cross-sectional Census results.

A number of robustness checks were performed. The age at migration was widened to age 20-to-54, but this did not alter the results (see Appendix Tables 7 and 8). Similarly, the sensitivity of the models to various positive earnings thresholds ($0; $1; $100; $1,000) and to trimming individuals at the top/bottom 1% of the distribution was also tested; neither had a significant impact.[17] Finally, regression models (1) and (2) were estimated on a pooled sample of all immigrant cohorts and all comparison groups. These models included period-of-arrival dummies and assume that the earnings trajectory is the same for all cohorts. Such estimations, while requiring an additional assumption, are quite common in the literature. The estimated coefficients from these regressions, reported in Appendix Tables 9 and 10, are consistent with the findings from the within-cohort regressions. That is, the longitudinal and cross-sectional estimates differ for absolute earnings growth but not for the earnings gap.

16. As in Lubotsky (2007), individuals' educational attainment is not controlled for, since the objective is to test for a bias in the unconditional earnings trajectories of immigrants and the native-born. The only difference with Lubotsky's specification is in the age variable, where he uses instead a potential experience variable (age minus years since completion of schooling). Education information for native-born is not available in Canada; consequently, for purposes of this paper, it is not possible to differentiate potential labour market experiences.
17. The results from these additional regressions are not reported, in the interest of brevity. They are available upon request.

Table 5
Relative immigrant earnings growth in Canada, longitudinal versus cross-sectional data

Years since migration	Change in earnings gap over time		
	LAD-IMDB		Census
	Longitudinal	Cross-sectional	
	Column 1	Column 2	Column 3
	coefficient from regression model		
1985-to-1989 cohort			
6 to 10 years	0.121	0.081	0.094
11 to 15 years	0.216	0.192	0.181
16 to 20 years	0.247	0.242	0.239
0 years (entry gap)	-0.348	-0.328	-0.334
1990-to-1994 cohort			
6 to 10 years	0.212	0.221	0.236
11 to 15 years	0.252	0.265	0.279
0 years (entry gap)	-0.447	-0.465	-0.444
1995-to-1999 cohort			
6 to 10 years	0.151	0.153	0.141
0 years (entry gap)	-0.267	-0.274	-0.224

Notes: The coefficients provide a rough estimate of percentage growth; e.g., the earnings growth in the longitudinal data between 0-to-5 years and 6-to-10 years in Canada for the cohort entering between 1985 and 1989 was approximately 12.1%. The reference category is immigrants with one to five years since migration. All coefficients are statistically significant at the 5% level. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

On the one hand, taken together, the empirical results of this study provide no evidence that cross-sectional estimates of immigrants' earnings relative to those of the Canadian-born are significantly biased by selective attrition, at least for the three immigrant cohorts included in the analysis. On the other hand, there is evidence that estimates of immigrants' absolute earnings growth based on repeated Census cross-sections are biased upwards slightly. This is likely due to a higher probability of exit from the sample by lower-earning immigrants. To better understand why there is a bias in the absolute earnings trajectory, but not in the earnings gap trajectory, the earnings characteristics of sample "exiters" and "stayers" are examined more closely.

5 Who is leaving the sample?

This section examines whether people who were in the cross-sectional administrative sample but not in the longitudinal sample tend to be disproportionately low earners rather than uniformly distributed throughout the earnings distribution. These people are referred to as "exiters." In other words: is the probability of exiting from the cross-sectional sample greater for low earners than for high earners? If it is greater for low earners, such an exit pattern would lead to an upward bias in earnings trajectories on the basis of cross-sectional data, of the kind observed above for absolute earnings growth but not for relative earnings.

To answer this question, it is necessary to define the earnings distribution and to determine where each "exiter" is located in the distribution, in the appropriate manner. The main issue to

be confronted is how to obtain the appropriate distribution for the entire immigrant cohort, given that some observations exit employment early in the study period, some exit later, and some remain to the end—i.e., the "stayers."

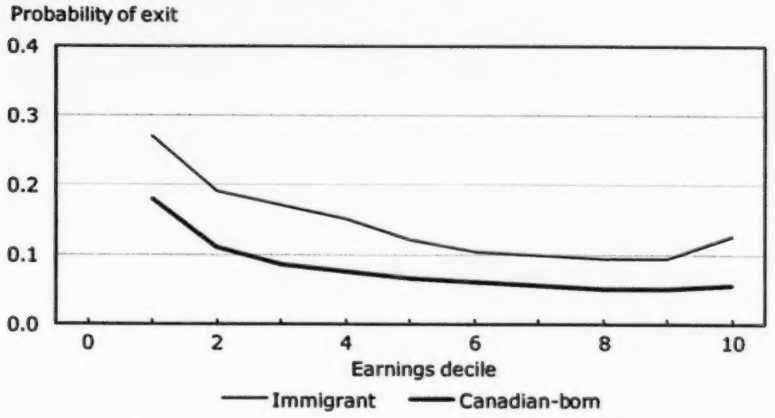
One could choose the earnings distribution that existed during the first few years in the host country. However, an immigrant's position in the distribution will likely change in later years as more time is spent in the host country. Consequently, such an approach could be misleading. A second option would be to compute mean earnings over the entire study period for each observation and to base the distribution on the individual means. However, earnings rise with years since migration. Hence, the earnings of early "exiters" would be significantly underestimated relative to those of stayers and of later "exiters," and their positioning in the earnings distribution would be incorrectly determined. What we want to know is how well the "exiters" did, economically, during the years they worked in Canada, relative to all other immigrants of the same cohort during exactly the same years in Canada. To achieve this, in any given year, *T*, a dummy variable *exit* is set to *1* for all individuals who left the sample in year *T* but who were observed in the sample in *T-1*. The cumulative earnings of each individual in year *T* is calculated as the sum of that person's earnings up to the previous year (*T-1*). This allows one to define a different earnings distribution for each year (e.g., for all years between 1986 and 2005 for the 1985 cohort) on the basis of the cumulative earnings between 1986 and any given year *T*. The earnings level of each "exiter" relative to that of all other immigrants in her/his cohort over exactly the same years can then be determined. This approach is applied to the year of first exit. Some "exiters" may have exited, re-entered, and subsequently re-exited the sample, but only the first exit is considered here.

A simple regression model is run for each earnings decile to estimate the probability of exit, *Pr(exit = 1)*. This is done by using a probit model where the dependent variable is the exit dummy as defined above. The independent variables include dummy variables for the 10 earnings deciles and a second-degree polynomial in age. The age controls are designed to eliminate any effect of the age-earnings profile on the likelihood of "exiters" being in the lower-earnings deciles.[18] The LAD-IMDB data are used for the purpose of this exercise for each of the three immigrant cohorts (1985, 1990, and 1995), as well as for each respective comparison group.[19]

Results are shown in Charts 1, 2 and 3. Each chart plots the probability of exit from the cross-sectional LAD-IMDB sample by earnings decile. Note that the probability of exit is always higher for immigrants than for Canadian-born. Some (or most) of this difference could be due to a higher probability of out-migration from Canada for immigrants, but the data do not allow one to differentiate the reasons for exiting the sample. When one examines immigrants only, the results show a much greater probability of exiting the sample for immigrants in the bottom three or four deciles than for immigrants in the top deciles. For the 1985 cohort, immigrants in the bottom decile had almost a 30% chance of exiting the sample, compared to around 10% for those in the top half of the distribution. A similar pattern is observed for the other two cohorts. This is the pattern leading to the upward bias in the absolute earnings of immigrants observed in the repeated cross-sectional data.[20]

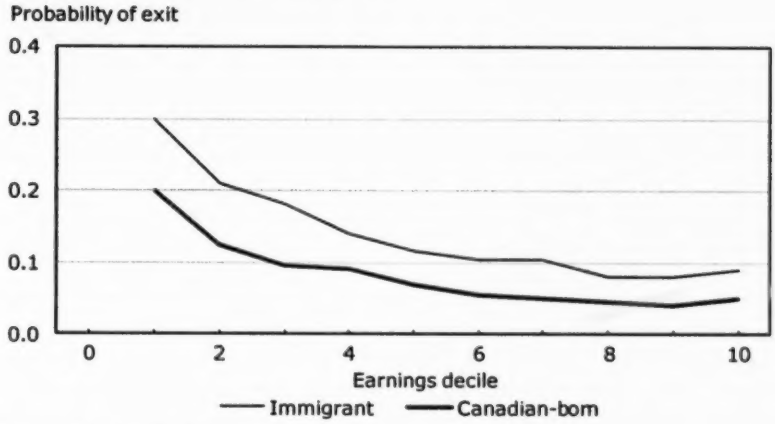
18. Young people tend to both have lower earnings and be more mobile than older workers.
19. This paper focuses on single-year cohorts in this section because doing so simplifies determining the cumulative earnings deciles. The findings generalize, however, to the other single-year cohorts.
20. An interesting pattern to note is the increase in the probability of exit at the very top of the earnings distribution for immigrants, particularly for the 1985 cohort. This could be related to the phenomenon of highly skilled immigrants using Canada as a stepping stone to the North American labour market.

Chart 1
Probability of exit by earnings decile — 1985 immigrant cohort and comparison group



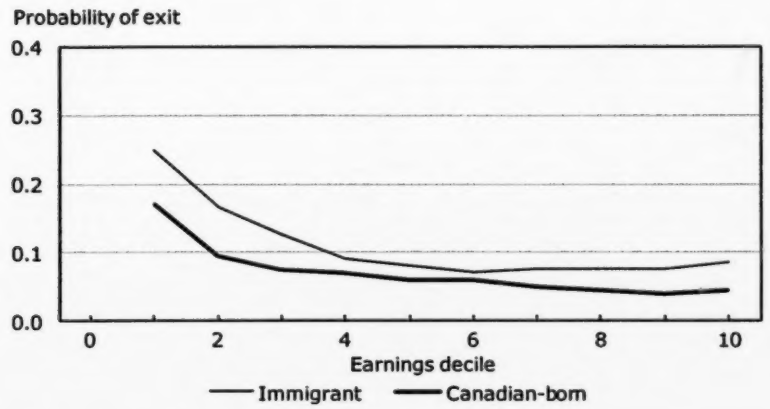
Note: Authors' estimation.
Sources: Statistics Canada, Longitudinal Administrative Database and Longitudinal Immigration Database.

Chart 2
Probability of exit by earnings decile — 1990 immigrant cohort and comparison group



Note: Authors' estimation.
Sources: Statistics Canada, Longitudinal Administrative Database and Longitudinal Immigration Database.

Chart 3
Probability of exit by earnings decile — 1995 immigrant cohort and comparison group



Note: Authors' estimation.
Sources: Statistics Canada, Longitudinal Administrative Database and Longitudinal Immigration Database.

The degree of the bias for any cohort would depend upon the difference in the probability of exit between low and high earners; this difference varies across cohorts. For the 1985 cohort, the probability of exit in the bottom three deciles (at around 20%) is 1.9 times that for the top three deciles (at around 11%). In the 1995 cohort, the ratio is 2.2. The differences are greatest for the 1990 cohort, where the probability of exit was 2.7 times greater among low earners (bottom three deciles) than among high earners. This is consistent with the greater upward bias in the absolute earnings trajectories in the 1995 cohort compared with the other cohorts estimated from the cross-sectional data.

But what of the fact that little bias was observed in the relative earnings trajectories in the cross-sectional data? When one focuses on the comparison group, virtually the same patterns are observed: the probability of exit is significantly higher among workers at the bottom of the earnings distribution than among those at the top. This explains why bias is not observed in the immigrant–Canadian-born earnings gap over time. The probability of exit is lower among Canadian-born than among immigrants at all points in the distribution. This is not surprising, as immigrants are more likely to separate from earnings than Canadian-born, particularly during their early years in Canada, when they experience lower job stability. However, the primary concern in this regard is with the relative probability of low and high earners exiting. Here, one can observe the same pattern among the Canadian-born as among immigrants. The earnings growth of both the immigrant cohorts and the comparison groups estimated from cross-sectional data is biased upwards as a result of the higher likelihood of lower-earning individuals exiting the sample. The gap, which is the difference between these two trajectories, is consequently only marginally affected, as the bias appears to apply to both immigrants and Canadian-born similarly.[21]

21. Results in this study do not vary when one uses a logit or linear probability model instead of a probit model. The findings are also robust to an alternative definition of cumulative earnings that does not count the earnings from the year prior to exit (in order to avoid potential issues deriving from part-year earnings data).

6 Conclusion

The majority of studies of immigrants' earnings trajectories have been based on Census data and have used pseudo-longitudinal cohorts constructed from repeated cross-sections. However, the population in these types of data changes over time, as some members of the immigrant cohort leave the country, or at least exit employment. If a selection effect is introduced by the less successful and lower-paid immigrants disproportionately exiting (as compared to higher earnings immigrants), the earnings growth estimated from repeated cross-sectional data will be biased upwards. Some of the earnings progress would not be "real" but rather would be due to the changing composition of the immigrant cohort over time.

In this study, longitudinal administrative data and Census data are used in order to determine whether such a bias exists in Canada. The focus is on both the absolute earnings trajectories of entering immigrant cohorts and the trajectories of the immigrant–Canadian-born earnings gap over their first 20 years in Canada. Regarding the growth in absolute earnings of immigrant cohorts, results based on repeated cross-section do differ from the results based on longitudinal data. In the three immigrant cohorts examined, the earnings growth (with years in Canada) based on a standard regression model was between 7-percentage-points higher and 15-percentage-points higher in the repeated cross-sectional sample than in the longitudinal sample. This bias was introduced because immigrants in these cohorts who were in the bottom one-third of the earnings distribution were at least twice as likely to exit the sample as those who were in the top one-third.

However, the change in immigrants' earnings *relative* to the earnings of the Canadian-born, which is the sole focus of most of the literature in this area, is very similar in the cross-sectional and longitudinal samples. Little evidence of a significant upward bias in the *earnings gap* trajectory is found. Although the less successful and lower-paid immigrants in the various cohorts are more likely to exit the sample, the same is true for the Canadian-born. That is, the earnings growth of both the immigrant and the Canadian-born samples is over-estimated in cross-sectional data, by roughly the same extent. Hence, the trajectory in the earnings gap estimated by means of longitudinal data and a standard earnings model does not point to bias in previous studies of immigrant earnings in Canada.

A comparison of these results to those in the United States (Lubotsky 2007) suggests that immigrants to Canada display labour market participation patterns that are more similar to those of the native-born than would appear to be the case in the United States.

7 Appendix

Table 6
Incidence of immigrants over time in the Census and the LAD-IMDB

Year	LAD-IMDB sample			Census sample			Census comparison group
	Sample size	Immigrants		Sample size	Immigrants		Immigrants who landed before 1980
	number		percent	number		percent	
1985-to-1989 cohort							
1990	801,772	19,494	2.43	863,459	20,746	2.40	5.98
1995	748,578	15,576	2.08	803,699	17,625	2.19	5.72
2000	710,157	14,689	2.07	749,061	16,645	2.22	5.72
2005	663,986	14,082	2.12	697,100	16,493	2.37	5.56
1990-to-1994 cohort							
1995	804,403	29,049	3.61	856,416	28,483	3.33	3.27
2000	776,789	26,703	3.44	815,563	26,326	3.23	3.31
2005	753,034	24,940	3.31	782,863	25,354	3.24	3.33
1995-to-1999 cohort							
2000	780,890	28,116	3.60	813,354	27,819	3.42	1.28
2005	770,008	24,925	3.24	795,307	25,643	3.22	1.34

Notes: Authors' calculations from the LAD-IMDB and the Census of Population. In each year, observations are restricted to individuals 25 to 64 years of age with positive earnings. Immigrants migrated between 25 and 44 years of age. *Comparison group* refers to the sample used as native-born group in the LAD-IMDB files (which includes immigrants landed before 1980). LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

Table 7
Sensitivity test for age at migration — Absolute immigrant earnings growth

Years since migration	LAD-IMDB		Census
	Longitudinal	Cross-sectional	
	Column 1	Column 2	Column 3
	coefficient from regression model		
1985-to-1989 cohort			
6 to10 years	0.103	0.056	0.036
11 to 15 years	0.251	0.255	0.236
16 to 20 years	0.248	0.330	0.311
1990-to-1994 cohort			
6 to 10 years	0.323	0.360	0.359
11 to 15 years	0.399	0.491	0.481
1995-to-1999 cohort			
6 to 10 years	0.206	0.258	0.233

Notes: The reference category is immigrants aged 20 to 54 at migration. All coefficients are statistically significant at the 1% level. The table reports the coefficients on the "years since migration" dummy variables in column 1. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

Table 8
Sensitivity test for age at migration — Relative immigrant earnings growth

Years since migration	Change in earnings gap over time		
	LAD-IMDB		Census
	Longitudinal	Cross-sectional	
	Column 1	Column 2	Column 3
	coefficient from regression model		
1985-to-1989 cohort			
6 to 10 years	0.111	0.084	0.082
11 to 15 years	0.195	0.187	0.181
16 to 20 years	0.215	0.222	0.223
0 years (entry gap)	-0.333	-0.323	-0.327
1990-to-1994 cohort			
6 to 11 years	0.201	0.216	0.218
11 to 15 years	0.247	0.265	0.276
0 years (entry gap)	-0.447	-0.469	-0.461
1995-to-1999 cohort			
6 to 10 years	0.148	0.151	0.144
0 years (entry gap)	-0.264	-0.266	-0.211

Notes: The reference category is immigrants aged 20 to 54 at migration. All coefficients are statistically significant at the 5% level. The table reports the coefficients on the "years since migration" dummy variables. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

Table 9
Absolute immigrant earnings growth across cohorts (pooled sample)

Years since migration	LAD-IMDB		Census
	Longitudinal	Cross-sectional	
	Column 1	Column 2	Column 3
	coefficient from regression model		
6 to 10 years	0.236	0.259	0.253
11 to 15 years	0.360	0.423	0.410
16 to 20 years	0.391	0.495	0.476

Notes: The reference category is immigrants with one to five years since migration. All coefficients are statistically significant at the 1% level. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.

Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

Table 10
Relative immigrant earnings growth across cohorts (pooled sample)

	Change in earnings gap over time		
	LAD-IMDB		Census
	Longitudinal	Cross-sectional	
	Column 1	Column 2	Column 3
	coefficient from regression model		
Years since migration			
6 to 10 years	0.171	0.164	0.168
11 to 15 years	0.234	0.233	0.232
16 to 20 years	0.264	0.277	0.269
Cohort dummies			
1985-to-1989 cohort	-0.323	-0.326	-0.332
1990-to-1994 cohort	-0.438	-0.446	-0.415
1995-to-1999 cohort	-0.329	-0.321	-0.285

Notes: Reference categories are immigrants with one to five years since migration. All coefficients are statistically significant at the 1% level. LAD: Longitudinal Administrative Database; IMDB: Longitudinal Immigration Database.
Sources: Statistics Canada, Longitudinal Administrative Database, Longitudinal Immigration Database and Census of Population.

References

Aydemir, A., and C. Robinson. 2008. "Global Labour Markets, Return, and Onward Migration." *Canadian Journal of Economics*. Vol. 41. No. 4. p. 1285–1311.

Baker, M., and D. Benjamin. 1994. "The performance of immigrants in the Canadian labor market." *Journal of Labor Economics*. Vol. 12. No. 3. p. 369–405.

Bloom, D.E., and M. Gunderson. 1991. "An analysis of the earnings of Canadian immigrants." *Immigration, Trade, and the Labor Market*. J.M. Abowd and R.B. Freeman (eds.). Chicago. The University of Chicago Press.

Borjas, G.J. 1999. "The Economic Analysis of Immigration." *Handbook of Labor Economics*. Vol. 3. D. Card and O. Ashenfelter (eds.). Amsterdam. North-Holland.

Chiswick, B.R. 1978. "The effect of Americanization on the earnings of foreign-born men." *Journal of Political Economy*. Vol. 86. No. 5. p 897–921.

Duleep, H.O., and D. Dowhan. 2002. "Insights from longitudinal data on the earnings growth of U.S. foreign-born men." *Demography*. Vol. 39. No. 3. p. 485–506.

Duleep, H.O., and M. Regets. 1997. "Measuring immigrant wage growth using matched CPS files." *Demography*. Vol. 34. No. 2. p. 239–249.

Edin, P.-A., R. LaLonde, and O. Åslund 2000. "Emigration of immigrants and measures of immigrant assimilation: Evidence from Sweden." *Swedish Economic Policy Review*. Vol. 7. p. 163–204.

Frenette, M., D.A. Green, and G. Picot. 2006. "Rising Income Inequality in the 1990s: An Exploration of Three Data Sources." *Dimensions of Inequality in Canada*. D.A. Green and J.R. Kesselman (eds.). Vancouver, British Columbia. University of British Columbia Press.

Frenette, M., and R. Morissette. 2003. *Will they ever converge? Earnings of immigrant and Canadian-born workers over the last two decades*. Statistics Canada Catalogue no. 11F0019M. Ottawa, Ontario. Analytical Studies Branch Research Paper Series. No. 215.

Grant, M.L. 1999. "Evidence of new immigrant assimilation in Canada." *Canadian Journal of Economics*. Vo. 32. No. 4. p. 930–955.

Green, D.A., and C. Worswick. 2004. *Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Macro Effects*. Mimeo.

Hu, W.-Y. 2000. "Immigrant Earnings Assimilation: Estimates from Longitudinal Data." *American Economic Review Papers and Proceedings*. Vol. 90. No. 2. p. 368–372.

Hum, D., and W. Simpson. 2000. "Closing the wage gap: Economic assimilation of Canadian immigrants reconsidered." *Journal of International Migrations and Integration*. Vol. 1. No. 4. 427–441.

Hum, D., and W. Simpson. 2004. "Reinterpreting the performance of immigrant wages from panel data." *Empirical Economics*. Vol. 29. No. 1. p. 129–147.

Lubotsky D. 2007. "Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings." *Journal of Political Economy*. Vol. 115. No. 5. p. 820–867.

McDonald, J.T., and C. Worswick. 1998. "The earnings of immigrant men in Canada: Job tenure, cohort, and macroeconomic conditions." *Industrial & Labor Relations Review*. Vol. 51. No. 3. p. 465–482.

Meng, R. 1987. "The earnings of Canadian immigrant and native-born males." *Applied Economics*. Vol. 19. No. 8. p. 1107–1119.

Picot, G., and A. Sweetman. 2005. *The Deteriorating Economic Welfare of Immigrants and Possible Causes*. Statistics Canada Catalogue no. 11F0019M. Ottawa, Ontario. Analytical Studies Branch Research Paper Series. No. 262.

Skuterud, M., and M. Su. 2009. *Immigrant Wage Assimilation and the Return to Foreign and Host-Country Sources of Human Capital*. Canadian Labour Market and Skills Researcher Network (CLSRN) Working Paper. No. 30.

Statistics Canada. 2009. *Longitudinal Administrative Data Dictionary, 2006*. Catalogue No. 12-585-X. Ottawa, Ontario.